International Journal of Electrical and
Electronics Engineering Research (IJEEER)
ISSN(P): 2250-155X; ISSN(E): 2278-943X
Vol. 6, Issue 1, Feb 2016, 1-12

TRANSFORMATION CODING FOR EMOTION SPEECH TRANSLATION: A REVIEW

P S DESHPANDE & J S CHITODE

Department of Electronics Bharati Vidyapeeth Deemed University College of Engineering, Pune, Maharashtra, India

ABSTRACT

This paper present a brief review on the current development for speech feature extraction and its usage in speech transformation for emotion speech translation. The process of prosody modeling and its application to speech transformation is reviewed in this paper. The past approaches of speech transformation from neutral to a emotion, for creating a synthetic speech signal is reviewed. The process of transformation and its relevant process of recognition, processing, and transformation is presented. A proposing system architecture for the transformation of a given neutral speech signal to its equivalent emotion speech is presented. The proposing architecture of speech transformation model is focused for its robustness in speech coding, transformation and accuracy in coding.

KEYWORDS: Speech Transformation, Emotion Coding, Prosody Modeling

Received: Dec 14, 2015; **Accepted:** Dec 21, 2015; **Published:** Jan 05, 2016; **Paper Id.:** IJEEERFEB20161

Review Article

INTRODUCTION

In the process of speech coding, speech transformation is an emerging area. Speech transformations have various significance of usage ranging from presentation of expression to forensic detection to crime investigation. The need of future speech synthesis lies on the not just detection but to validate its content and apply to various critical applications. The medical diagnosis, authentication usage, cinematography are few such applications. Wherein speech are an effective way of recognizing user, the spoken format rely its significance. Speech with its emotion reveals the personal objective of the spoken speech. It is hence required not only to know the speech spoken but also the emotion at which it is spoken as. The process of emotion synthesis was performed using prosody modeling. Wherein prosody model are used to detect spoken speech emotion, it can also be used to generate synthetic speech signal from given speech input. However, the processing accuracy of the transformation from one emotion to other is dependent on the prosody features. Hence it is required to derive the prosody model more accurately so as to achieve proper transformation. Wherein various approaches were developed to achieve the objective of proper extraction of prosody feature information from the speech signal, and mapping for synthetic transformation, the effect of processing noise, environmental distortion or magnitude variations lead to wrong transformation of the speech signal. Hence it is required to have a accurate feature extraction and transformation for emotion synthesis in speech coding. To observe the performance of the proposing approach, the proposing approach is to be evaluated over variant conditions such as noises, magnitude variation, gender variation, age variation for different emotion transformation. Towards such transformation and feature extraction, various developments were made. The approach of feature extraction and its usage to transformation is summarized in the following section.

STATE OF THE ART

Wherein the process of prosody detection and processing for speech coding is based on the extraction of domain speech features and transforming to a given emotion, the process of recognition, classification, and feature extraction are prime importance towards its processing accuracy. Towards the recognition of emotion or its transformation various developments were observed.

- **Emotion Recognition**

For the development of recognition of speech prosody for speech transformation in [1] a review towards expressive speech synthesis is presented. The review outlines the base models for speech transformation for expressive speech transformation (ESS) for human –computer interaction (HCI). In [2] an animated format of the prosody feature processing for speech transformation is presented. For a audio-visual representation for expressive speech communication. A realistic speech behavior for speech coding in this application was presented. In [3] a multi lingual speech transformation based on prosodic feature to develop a speech-to-speech translation. The prosodic features are derived using unsupervised clustering algorithm and mapped for the target speech transformation based on the targeted speech quality. In [4,5], a cross-modal priming task was employed to perform the categorization of emotions from a speech signal. In this approach, after listening to angry, sad, disgusted, or neutral vocal primes, subjects rendered a facial affect decision about an emotionally congruent or incongruent face target. In [6], a study was performed to investigate the effects of emotion class, sentence type on the speech amplitude and speaker. This approach also studied the recording technique to maintain dynamic information for a complete full blown speech. The results of this approach revealed that the speaker and emotion class are highly significant and later reveals the half of variance. In [7], a hybrid approach was proposed by combining the template parametric manipulation with prosody parametric manipulation for the purpose of quality improving and also to generate prosody for boeo speech. This approach also aimed to increase the annotation variability of synthesized speech output. This approach also considered prosodic features for emotion detection from speech signal. The extracted prosody is the combination of intensity, pitch and duration. This approach considered anger, happiness, fear and sadness for synthesis evaluation. Through this evaluation this approach found the accurate prosody of bodo speech to confirm perception tests. In [8], a short introduction was given about speech emotion recognition. This also gives the information about prosody features, speech emotion and speech style and also more other information was also involved. The main uncertainties of speech recognition were also outlined in this. In [9], a novel network called as probabilistic echo state network (pi-SEN) was proposed to find the density over a variable having length sequences and also multivariate domain vectors. This pi-ESN was formed by combining the parametric density based radial basis function and reservoir fan ESN. At classification stage, this approach used maximum likelihood training process. The feature parameters used for prosody representation are effective to such recognition process.

- **Feature Description**

In [10] a time domain analysis for speech transformation based on given input signal is presented. The pitch modification approach was presented with the process of time based coding for speech signal using linear prediction logic. The linear prediction residual is used as a pitch variant parameter for speech transformation. The process of linear prediction also helps in altering the shape content of the speech signal. In [11] new quality features namely formants, spectral energy distribution in different frequency bands, harmonics-to-noise ratio (in different frequency bands) and irregularities (jitter, shimmer). The process of Dimensional approach for emotion classify is carried out. It is shown that

these quality features are more appropriate in representation in comparison to different valance levels in dimensional approach. In [12], an approach was proposed to study the effect of emotional speech prosody on the participants and fixate features that congruent of an emotional speech of prosody. In this approach, totally, 21 participants are observed for face expression such sadness, fear, happiness while listening to en emotionally uttered utterance spoken in a incongruent or congruent prosody. The all participants tried to judge the meaning of the voice or face of emotion whether it is same or not. In this study, it was confirmed that the eye movements will play an important role to match this. In [13,14], a novel approach is being proposed with a combination of prosody features (i.e. energy, pitch and Zero crossing rate), derived features (i.e. Linear Predictive Coding Coefficients (LPCC), Mel-Frequency Cepstral Coefficient (MFCC)),quality features (i.e .Spectral features, Formant Frequencies), and dynamic feature (Mel-Energy spectrum dynamic Coefficients (MEDC)) for an efficient automatic recognition of speaker's state of emotion. MSVM was used at classifier stage to classify the emotions such as happy, neutral, fear, sad and anger for a given speech signal. In [15], an approach was proposed to find seven types of emotions of a speech signal. They are, fear, anger, sadness, boredom, neutral, disgust and happiness. Various DWT decompositions are used for feature extraction. SVM was used for classification. In [16] a emotion detecting framework was focused to design that performs various actions like feature extraction, speech to text conversion, feature selection and feature classification those are required for emotion detection. This approach uses prosody feature for emotion detection. The classification involves the training of various models and testing of a query sample. This approach extracts the features such that, they will convey measurable level of emotional modulation.

- **Classification and Coding**

In [17,18], human speech emotion recognition system was developed based on the acoustic features like pitch, energy etc, and spectral feature MFCC. Then SVM and CART has been used as classifier. The complete implementation of proposed approach was done under two phases: training and testing. This approach was tested over different voice files of different emotions like Anger, Disgust, Fear, Happy, Neutral, Sad and Surprise. [19] Presented the approach of speech transformation using hidden markov model, for phonotic unit detection corresponding to the training data set. The model uses the dynamic property of HMM for speech recognition. A MFCC based feature extraction is used for the speech feature detection and HMM modeling is used toward decision deriving for emotion detection. [20] presents an approach for text –to-speech transformation using prosody model in context to the tone languages. A fuzzy based modeling for speech coding is proposed. A classification and regression model tree (CART) based on testing and modeling of speech duration is proposed. A tree modeling for speech detection to observe the syllable in word and its position in sentence was made. The length of the word, peak of the targeted syllable, the phonetic structure of the syllable were also derived for speech coding. In [21,22] a Gaussian mixture model (GMM) based on emotional voice transformation using the conversion of prosody feature was proposed. The GMM based modeling is used for the conversion of non-linguistic information, while keeping the linguistic data intact. The objective of such conversion was to keep the prosody processing with greater level of speech quality. An approach of neural network based speech synthesis system was presented in [23]. The approach of neural network is carried out for the prosodic feature of the syllables for their position, context and phonological feature. The prosody model developed were evaluated for the text to speech transformation, used for speech recognition, speech synthesis and speaker recognition. The NN model is developed for the capturing of dialects in Hindi language. Towards database optimization in [24,25,26,27,28] an approach to minimize the dataset collection and processing effort were developed focusing on maintaining acceptable quality and naturalness for speech transformation. A text-to-speech (TTS) framework MARY was developed with voice quality using GMM-based prediction and vocal tract processing. In [29,30],

an approach made an attempt to synthesize emotional speech by using “strong,”“medium,” and “weak” classifications. In this approach, various linear models were processed for test like Gaussian Mixture Model (GMM), Linear Modification Model (LMM) and a classification and Regression Tree (CART). This paper also analyzes objective and subjective analyses. [31] Focused on happiness. This paper computes the emotional speech for happiness for a given speech signal. In this approach, totally 11 types of emotional utterances were developed; each and every utterance was labeled with a PAD value. This paper also proposed a five-scale tone model to model the contour of the pitch. A generalized regression neural network (GRNN) situated prosody conversion model is constructed to recognize the transformation of pitch contour, duration and pause duration of emotional utterance, where the PAD values of emotion and context parameter are adopted to predict the prosodic aspects. Emotional utterance is then re-synthesized with the STRAIGHT algorithm by way of modifying pitch contour, length and pause period. In [32,33,34], a novel speech synthesis approach was proposed using the subspace constraint. This approach employs Principal Component Analysis (PCA) to reduce the prosodic components dimensions. Then the obtained samples of PCA are trained. The features included in this approach are mainly F0, speech length, power and finally the correlative length. This approach also assumed that the combination of accent type and syllables will determine the correlative dynamics of prosody. This approach successfully worked for synthesized speeches especially, “disgust”, “boredom”, joy“, “anger”, “surprise”, “sorrow” and “depression”. In [35], an approach was proposed with the aim of speech enhancement and also with the aim of hearing impaired patients. This approach didn’t require more training data. This approach also studies the conversion of prosodic samples. This approach used Eigen voice- Gaussian Mixture Model (EV-GMM) to transform the spectral parameters and F0. [36] Investigated the effect of musical train on the observation of vocally expressed emotion. This approach takes the base of Event-related Potential correlates for emotional prosody processing. 14 control subjects and 14 emotional musicians are listened to 228 sentences with intelligible semantic content, neutral semantic content, differing in prosody and unintelligible content. This investigation observes that the P50 amplitude was reduced. A difference between SCC and percent stipulations used to be discovered in P50 and N100 amplitude in non-musicians only, and in P200 amplitude in musicians simplest. [37,38]represents of micro intonation and spectral characteristics in female and male acted emotional speech.Micro intonation element of speech melody is analyzed related to its spectral and statistical characteristics. Consistent with psychological study of emotional speech, exclusive feelings are accompanied by different spectral noise. We manipulate its amount by using spectral flatness in step with which the excessive frequency noise is mixed in voiced frames throughout cepstral speech synthesis. In [39], an emotional speech synthesis framework was proposed to analyze the effect of emotions in the story telling speech. This approach used XML and SABLE languages to synthesize the emotions of text. The SABLE language was used for the purpose of speech quality improvement from the contaminative speech synthesizer. This approach used a set of tags of prosody to synthesize the emotion of the speech from a given text. The prosody correlates of pitch range, pith base and intensity was found by Modified Zero frequency filtered (ZFF) signal. Then the required prosody parameters are stored in a template format. At synthesis stage, prosody tags were replaced by hand annotated text story. The naturalness and quality of the synthesized emotional speech was analyzed through subjective tests. In [40], a new approach was proposed to find the speech emotion recognition through the Fuzzy Least squares and Support Vector Machines (FLSSVM). This FLSSVM constructs the optimal hyper plane using the feature extracted from the voice and prosody, recognizes four main emotions sadness, happiness, anger and surprise respectively. Compared with earlier approaches this approach performs efficient emotion recognition. [41] Proposed a new emotional recognition approach with the aim of three main goals. The first one is to update the up to date available emotion speech data collections. Next, the second goal is to extract the features efficiently such that they are able to find emotion more accurately and also to measure their effect. This approach considered the

features asvocal tract cross-section areas, Teager energy operator-based features, pitch, the mel-frequency cepstral coefficients and the speech rate. The third and last goal is to analyze the techniques for the classification of emotional states. These techniques include, HMM model, ANN mode, LDA model, K-NN model and finally SVM model. In [42,43], a speech recognition and understanding approach was proposed to model the super segmental characteristics of speech and also for acoustic processing by including the advanced semantic and syntactic level processing. The proposed approach was completely based on HMM modeling. The HMM is used in this paper to model the speech prosody and to make the initial semantic level processing of input speech. The energy and fundamental frequency were used as acoustic features. An approach was also applied for semantic features extraction. The method was designed to work for fixed-stress languages, and it yields a segmentation of theinput speech for syntactically linked word agencies, or even single phrases akin to a syntactic unit (these word companies are regularly referred to as phonological phrases in psycholinguistics, which can include a number of words). These so-referred to as phrase-stress items are marked by means of prosody, and have an associated fundamental frequency and/or energy contour which enables their discovery. This semantic level processing of speech was investigated for the Hungarian and for the German languages. In [44], the prosodic features such as intonation, duration and intensity patterns were studied. The basic difference between neutral speech and emotional speech is prosodic features only. In this approach, three acoustic features are studied: duration, intensity and pitch contour. This study was imposed on Hindi emotional speech and on Hindi neutral speech. The praat tool was used to impose acoustic features on neutral speech. The medical perspective of the speech generation and the brain state condition reflects the prosodical feature representation of the speech emotion.

- **Medical Processing**

In [45] an application of speech prosody, a medical diagnosis of unpaired discrimination for different speech emotion were presented. The effect of speech emotion on the detection of medical disorderness is outlined. The presented task suggests the usage of speech synthesis on fearness detection. In [46,47] a generation of expressions on speech coding based on human neural analysis is presented. This approach presents a unique approach of neural based emotional prosody analysis in human speech coding. The bilateral operation of human brain on the prosody modeling and its operations are illustrated. In [48,49,50] to evaluate the effect of schizophrenia in patients for prosody future extraction, a feature extraction process for emotion feature extraction process is carried out. 94 patient and 51 healthy subjects were taken for the evaluation of the suggested approach. The emotion perception disorderness due to the disease patient was analyzed. In [51,52], electrophysiological and behavioral measurements based word processing was done to study the effect of sad prosody in hemispheric specialization. By combining the signal detection method with focused attention, a dichotic listening method was proposed in this approach to find the words spoken in sad prosody or neutral prosody. This approach finalizes that the single sad prosody is not able to modulate the asymmetry of hemisphere in word level processing. In [53], an investigation was done towards the hemispheric contributions for emotional speech processing by comparing adults with a focal lesion of left and right hemisphere and also with brain damage. Participants listened to semantically anomalous utterances in three conditions (discrimination, identification, and score) which assessed their attention of 5 prosodic feelings beneath the effect of one of a kind mission- and response-choice demands. Findings printed that correct- and left-hemispheric lesions had been associated with impaired comprehension of prosody, even though possibly for distinct motives: correct-hemisphere compromise produced a more pervasive insensitivity to emotive points of prosodic stimuli, whereas left-hemisphere harm yielded better difficulties decoding prosodic representations as a code embedded with language content. [54,55] Used event-related brain potentials (ERPs) to evaluate the time path of emotion processing from

non-linguistic vocalizations versus speech prosody, to scan whether vocalizations are dealt with preferentially via the neuro cognitive method. Members passively listened to vocalizations or pseudo-utterances conveying anger, sadness, or happiness because the EEG used to be recorded. Simultaneous effects of vocal expression style and emotion were analyzed for 3 ERP add-ons (N100, P200, late confident factor). Emotional state of a speaker is accompanied with the aid of physiological alterations affecting breathing, phonation, and articulation. These changes are manifested on the whole in prosodic patterns of F0, energy, and duration, but also in segmental characteristics of speech spectrum. For that reason, a new emotional speech synthesis process proposed in [56] is supplemented with spectrum amendment. It includes non-linear frequency scale transformation of speech spectral envelope, filtering for emphasizing low or excessive frequency range, and controlling of spectral noise via spectral flatness measure in keeping with skills of psychological and phonetic study. The Aprosodia Battery was once developed to distinguish unique patterns of affective-prosodic deficits in sufferers with left versus correct mind injury by way of using affective utterances with incrementally reduced verbal-articulatory demands. [57,58] describes an huge, quantitative error evaluation utilizing previous outcome from the Aprosodia Battery in patients with left and right brain harm, age-identical controls (historical adults), and a gaggle of younger adults. This inductive evaluation was performed to deal with three foremost disorders in the literature: (1) sex and (2) maturational-getting older results in comprehending affective prosody and (3) differential hemispheric lateralization of emotions. This approach found no overall sex effects for comprehension of affective prosody. There have been, nonetheless, scattered sex results involving a specific have an impact on, suggesting that these variations have been concerning cognitive appraisal alternatively than most important perception. In [59,60], an emotion recognition approach was proposed especially for children with HFA (n=26, 6-11 years) based the prosody features of them. In this approach, Vineland adaptive behavior scale and communication checklist of children was used to assess social and pragmatic abilities.

- **Language Coding**

In [61] a speech processing with pause factor is considered. The development reveals the effect of speech pause on the recognition process. The process is carried over Hungarian and Austrian data set, manipulating the emotional conversion. In [62,63] towards a speech transformation based on the effect of emotion transformation in Hindi speech processing a Hindi speech database is developed. Pitch feature were derived for different emotion and used as a transformation metric for speech transformation. Three tests transformation for sad, joy and anger were developed for the transformation from neutral speech signal. In [64] a speech transformation process for kannada language was developed. An approach of linear modification model (LMM) was used. This method is used for the conversion of emotion data to its target emotion. A kannada dataset is been created and the effect of change in emotion is evaluated. The effect of pitch on the speech transformation is carried out over the pitch points extracted. The process was carried out over sadness and fear cases. In [65,66] an approach to speech synthesis for Czech and Slovak emotion speech based on spectral coding for prosodic feature is developed. A Gaussian mixture model (GMM) is used for processing speech classification using GMM training process. The selection process of feature and its impact on emotion classification were analyzed. The functionality of a 2 level architecture comprising of gender detection and classification is also developed for speech transformation. The length of the speech signal were varied for different dimension and the effect of speech transformation were observed. The classification accuracy for the developed system is evaluated for detecting the effectiveness of the speech transformation approach. In the process of speech translation a speech synthesis model for text-to-speech for Marathi language was presented in [67,68,69]. The pitch features were focused based on its magnitude , count and contour used for transformation. The pitch factor is derived for speech with punctuation marks and processed to derive prosodic features.

The approach of neutral Marathi speech to emotional speech is presented based on pitch modification and word detection is presented. In [70], Hindi speech corpus signals have been used for simulated emotion detection. The data base was collected from the Gyanavani FM radio station, Varanasi, India of professional artists. The complete speech corpus was collected for eight emotions, are anger, disgust, fear, happy, neutral, sad, sarcastic and surprise. Emotion classification is performed on the proposed corpus using prosodic and spectral features. Energy, pitch and duration are used to represent prosody information. Mel frequency cepstral coefficients (MFCCs) are used to represent spectral information. [71] Proposed an approach to perform the German text to speech synthesis. This approach mainly used XML for internal data representation of system data, modular design. This approach also provides an interface for users to modify intermediate processing steps without any need of understanding of system technically. This approach also provides a detailed description through examples. In [72], a practical approach was proposed to generate the speech of Punjabi language. In Punjabi language, there are so many discontinuities. This work proposed to increase the utterances of speech by overcoming the problem of discontinuity in order to increase the naturalness. This approach was accomplished to increase the quality, natural resound of speech and simulated over various Punjabi audio files. In [73], a new prosody rule set was designed to convert the neutral speech to storytelling speech in Hindi language. This approach considered the features as pitch, duration, intensity and tempo to perform the conversion. For each and every prosodic parameters mentioned above, specific prosodic rules were built to denote the story teller emotions such as anger, sad, fear, neutral and surprise. For this purpose, some professional story tellers were gone to consult. The complete rules are derived both from male and female speakers. With the developments observed in past works, the speech synthesis for transformation are basically been carried out, using prosody parameter extraction and mapping. Wherein most of the works are focused towards development of transformation approaches for feature extraction, emotion recognition, and its transformation, less concentration is given towards the data content. During the process of speech coding it is observed that system noises such as hiss sound, or magnitude difference due to difference in capturing elements or variation in speech data varying with sampling rates, period etc. Though these content impact the speech transformation process less focus is made in this area. With this objective, a system design for robust speech transformation logic is proposed, as outlined in next section.

SYSTEM OUTLINE

To develop a robust transformation system following prosodic features for emotion transformation, system architecture is outlined. Figure 1 illustrates the block diagram for the proposed speech transformation approach.

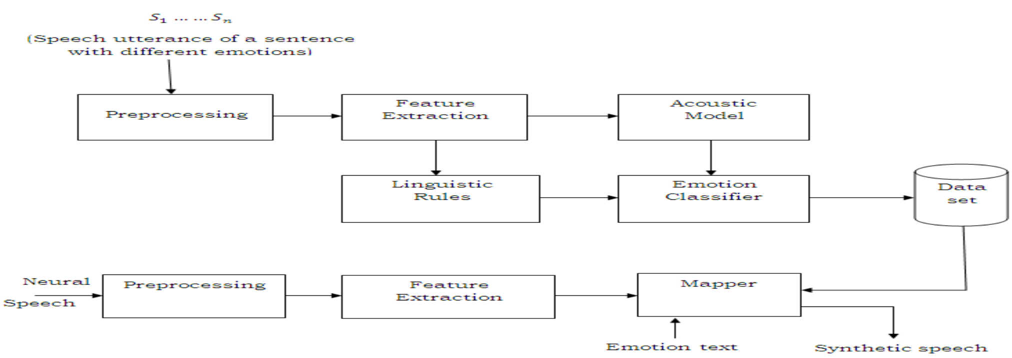

Figure 1: System Architecture for Proposed Emotion Speech Transformation

The process of speech transformation is carried out in two phases, of training and testing. In the process of training, a set of speech utterance for multiple subjects with different emotion will be captured. These samples are recorded

with maximum standard environments to obtain highest degree of speech clarity in recorded speech. These samples will then be passed for preprocessing, wherein the samples will be linearized to a uniform sampling rate, period and filtration. The windowing technique will be applied and energy spectral coding will be developed towards filtration process. To the processed speech signal, prosody features will be extracted, namely the jitter, shimmer, Harmonic to noise ratio (HNR), degree of voice breaks (DVB), degree of voiceless (DUV), pitch mean, pitch contour, pitch peaks, and pitch variance. These extractions of feature will be performed over all the training sample and a acoustic model with voice quality feature is developed. The linguistic rules will be extracted from the speech signal, for pause, breaks, and pronounce for each user. These linguistic rules are then passed to a emotion classifier where the emotion classification is performed based on SVM approach. The classified emotions with their corresponding features are then recorded to formulate feature dataset. During the test process, a neutral sample is passed, which is processed with the same process as carried out during training. The pre-processed sample is then processed for feature extraction and passed to a mapper logic where in based on the text input from the user prosodic feature will be derived from the dataset. The mapper logic transforms the neutral feature to the demanded emotion feature by magnitudal and spectral alignment to obtain the transformed synthetic speech.

CONCLUSIONS

This paper presented a brief literature outline in the area of prosody model for emotion speech transformation. The speech coding for emotion recognition and its representation is presented. The dominating extracting features for speech signals, which has higher impact on the speech transformation are derived. Most of the transformation or classifying process are performed using HMM. Wherein advance intelligence logic such as neural network and fuzzy logics were also used in few context. However as observed, less emphasis is given on speech content for transformation process. Though these factors also affects the transformation quality, in consideration to this a system architecture is proposed for robust emotion speech transformation.

REFERENCES

1. *Jasmine Kaur, Parminder Singh, "Review on Expressive Speech Synthesis", International Journal of Electrical, Electronics and Computer Systems, Vol.3, Issue-10, 2015.*
2. *Bjorn Granstrom , David House, "Audiovisual representation of prosody in expressive speech communication", Speech Communication, Vol.46, Elsevier, 2005.*
3. *Pablo Daniel Aguero, Jordi Adell and Antonio Bonafonte, "prosody generation for speech-to-speech translation", Acoustics, speech and signal processing, ICASSP, international conference IEEE, 2006.*
4. *Abhishek Jaywant, Marc D. Pell, "Categorical processing of negative emotions from speech prosody", speech communication, Vol.54, Elsevier, 2012.*
5. *M.B.Chandak Dr. Rajiv Dharaskar, "Emotion Extractor: A methodology to implement prosody features in Speech Synthesis", electronic computer technology (ICECT), international conference IEEE, 2010.*
6. *Hartmut R. P.tzinger, "Amplitude and Amplitude Variation of Emotional Speech", Inter Speech, 2008.*
7. *Laba kr. Thakuria, "Integrating Rule and Template- based Approaches to Prosody Generation for Emotional BODO Speech Synthesis", Fourth International Conference on Communication Systems and Network Technologies, IEEE, 2014.*
8. *Klara Vicsi, "Thinking about present and future of the complex speech recognition", 3rd International Conference on Cognitive Info communications, IEEE, 2012.*

9. *Edmondo Trentin, "Emotion recognition from speech signals via aprobabilistic echo-statenet work", pattern recogntion, Elsevier 2014.*
10. *Joao P. Cabral and Luıs C. Oliveira Pitch-Synchronous Time-Scaling for Prosodic and Voice Quality Transformations, Inter Speech 2005.*
11. *Martin Borchert and Antje Diisterhoft, "Emotions in Speech - Experiments with Prosody and Quality Features in Speech for Use in Categorical and Dimensional Emotion Recognition Environments" Proceeding ofNLP-KE'05. Proceedings of IEEE international conference on Oct. 2005.*
12. *Simon Rigoulot, "Emotion in the voice influences the way we scan emotional faces", speech communication, Vol.65,Elsevier 2014.*
13. *Amiya Kumar Samantaray, "A novel approach of Speech Emotion Recognition with prosody, quality and derived features using SVM classifier for a class of North-Eastern Languages", 2nd International Conference on Recent Trends in Information Systems, IEEE, 2015.*
14. *Rahul. B. Lanjewar, D. S. Chaudhari, "Speech Emotion Recognition: A Review", International Journal of Innovative Technology and Exploring Engineering, Vol.2, Issue-4, 2013.*
15. *Simina Emerich Eugen Lupu, "Improving Speech Emotion Recognition Using Frequency and Time Domain Acoustic Features", Proceedings of SPAMEC, Cluj-Napoca, Romania EURASIP 2011.*
16. *Rode Snehal Sudhkar, "Analysis of Speech Features for Emotion Detection: A review", International Conference on Computing Communication Control and Automation, IEEE 2015.*
17. *Anand C1, Devendran, "Speech Emotion Recognition using CART algorithm", International Research Journal of Engineering and Technology, Vol.02, Issue,03, 2015.*
18. *Prashant Aher, Alice Cheeran, "Comparative Analysis of Speech Features for Speech Emotion Recognition", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, 2014.*
19. *V.V. Nanavare, S.K.Jagtap, "Recognition of Human Emotions from Speech Processing", Procedia Computer Science 49, Elsevier, 2015.*
20. *odetunji A. odejobi, Shun Ha Sylvia Wong a, Anthony J. Beaumont, "A modular holistic approach to prosody modeling for Standard Yoruba speech synthesis", Computer Speech and Language 22 39–68 ELSEVIER 2008.*
21. *Ryo Aihara, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Ariki, "GMM-Based Emotional Voice Conversion Using Spectrum and Prosody Features", American Journal of Signal Processing, vol. 2, Issue.5, 2012.*
22. *Chung-Hsien Wu, Chi-Chun Hsia, Chung-Han Lee, and Mai-Chun Lin, "Hierarchical Prosody Conversion Using Regression-Based Clustering for Emotional Speech Synthesis", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 6, August 2010.*
23. *K. Sreenivasa Rao, "Role of neural network models for developing speech systems", Sadhana, Vol. 36, Part 5, Indian Academy of Sciences, October 2011.*
24. *Oytun Turk and Marc Schroder, "Evaluation of Expressive Speech Synthesis With Voice Conversion and Copy Resynthesis Techniques", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 5, July 2010.*
25. *Joe Crumpton Cindy L. Bethel, "Validation of Vocal Prosody Modifications to Communicate Emotion in Robot Speech", collaboration technologies and systems (CTS), international conference IEEE 2015.*

26. *K.C. Rajeswari, G. Karthick Prabu, "Developing Intonation Pattern for Tamil Text To Speech Synthesis System", international Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Issue 1, 2014.*
27. *P. Gangamohan, V. K. Mittal and B. Yegnanarayana, "A Flexible Analysis Synthesis Tool (FAST) for studying the characteristic features of emotion in speech", 9th Annual IEEE Consumer Communications and Networking Conference - Special Session Affective Computing for Future Consumer Electronics, IEEE, 2012.*
28. *J Hirschberg, "Speech Synthesis: Prosody", Elsevier 2006.*
29. *Miss Ashbin S Shinde ,Mr Sachin S Patil, "Emotion classification and frequency domain parameters of speech signal for the expression of prosody in synthetic speech", IOSR Journal of Electrical and Electronics Engineering , 2014.*
30. *Jianhua Tao, Yongguo Kang, and Aijun Li Prosody, "Conversion From Neutral Speech to Emotional Speech", IEEE Transactions on audio, Speech, and Language Processing, Vol. 14, No. 4, July 2006*
31. *Xiaoyong Lu, Hongwu Yang, Aibao Zhou, "Applying PAD Three Dimensional Emotion Model to Convert Prosody of Emotional Speech", Orange Technologies (ICOT), international conference IEEE 2014.*
32. *Juan Pablo Arias, Carlos Busso, Nestor Becerra Yomaa, "Shape-based modeling of the fundamental frequency contour for emotion detection in speech", Computer Speech and Language 28 278–294, Elsevier 2014.*
33. *Shinya Mori, Tsuyoshi Moriyama, Shinji Ozawa, "Emotional Speech Synthesis Using Subspace Constraints In Prosody", Multimedia and Expo, international conference IEEE 2006.*
34. *Soumaya Gharsellaoui, Sid-Ahmed Selouani, Adel Omar Dahmane, "Automatic Emotion Recognition using Auditory and Prosodic Indicative Features", 28th Canadian Conference on Electrical and Computer Engineering, IEEE, 2015.*
35. *Shulan Xia, "A speech emotion enhancement method for hearing aid", Computer Modeling & New Technologies Vol.18, Issue 11, 2014.*
36. *Ana P. Pinheiro, "The music of language: An ERP investigation of the effects of musical training on emotional prosody processing", Brain and Language, Elsevier, 2015.*
37. *J. Pribil "Statistical Analysis of Spectral Properties and Prosodic Parameters of Emotional Speech", Measurement Science Review, Volume 9, No. 4, 2009.*
38. *Kurt Hammerschmidt and Uwe Jurgens, "Acoustical Correlates of Affective", Prosody Journal of Voice, Vol. 21, No. 5, 2007.*
39. *Gurunath Reddy M, Harikrishna D M, K. SreenivasaRao and Manjunath K E, "Telugu Emotional Story Speech Synthesis using SABLE Markup Language", SPACES, Dept of ECE, K L University, 2015.*
40. *Shiqing Zhang, "Speech Emotion Recognition Based on Fuzzy Least Squares Support Vector Machines", 7thWorld Congress on Intelligent Control and Automation, 2008.*
41. *Dimitrios Ververidis, "Emotional speech recognition: Resources, features, and methods", Speech communication, Elsevier 2006.*
42. *Klara Vicsi, "Using prosody to improve automatic speech recognition", Speech communication, Elsevier 2010.*
43. *Ishpreet Kaur, Manveen Kaur, Oberoi Simrat Kaur, "Prosody Modification of Recorded Speech in Time-Domain", International Journal of Computer Application , 2012.*
44. *Jainath Yadav, "Generation of emotional speech by prosody imposition on Sentence, Word and Syllable level fragments of neutral speech", Cognitive computing and information processing (CCIP), international conference, IEEE 2015.*
45. *Dominik R.Bach, René Hurlemann, Raymond J.Dolan, "Unimpaired discrimination of fearful prosody after amygdala lesion",*

Neuropsychologia, Elsevier, Vol. 51, 2013.

46. *Swann Pichon, and Christian A. Kell, "Affective and Sensorimotor Components of Emotional Prosody Generation", The Journal of Neuroscience, Vol.33, 2013.*
47. *Thomas Ethofer, Benjamin Kreifelts, Sarah Wiethoff, Jonathan Wolf, Wolfgang Grodd, Patrik Vuilleumier, and Dirk Wildgruber, "Differential Influences of Emotion, Task, and Novelty on Brain Regions Underlying the Processing of Speech Melody", Journal of Cognitive Neuroscience Vol. 21, issue 7, 2008.*
48. *Filomena Castagna, CristianaMontemagni, AnnaMariaMilani, GiuseppeRocca, PaolaRocca, Massimo Casacchia, FilippoBogetto, "A Prosody recognition and audiovisual emotion matching in schizophrenia: The contribution of cognition and psychopathology", Psychiatry Research, 2013.*
49. *Jia Huanga, Raymond C.K. Chana, Xiaobin Lue, Zheng Maf, Zhanjiang Lif, Qi-yong Gongg, "An exploratory study of the influence of conversation prosody on emotion and intention identification in schizophrenia", Brain Research, Vol. 58, Elsevier 2008.*
50. *Jaimi Marie Iredale, Jacqueline A. Rushby, Skye McDonald, Aneta Dimoska-Di Marco, "Joshua Swift Emotion in voice matters: Neural correlates of emotional prosody perception", International Journal of Psychophysiology Vol.89, Elsevier 2013.*
51. *RotemLeshem, "The effects of sad prosody on hemispheric specialization for words processing", Elsveir, 2015.*
52. *Konstanty Guranski , Ryszard Podemski , "Emotional prosody expression in acoustic analysis in patients with right hemisphere ischemic stroke", neurologia and neurochirurgia polska Vol.49, Elsevier 2015.*
53. *Marc D. Pell, "Cerebral mechanisms for understanding emotional prosody in speech", Brain and Language Elsevier 2006.*
54. *M.D. Pell, "Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody", Biological Psychology, Elsevier 2015.*
55. *Yingying Gao, Weibin Zhu, "Detecting affective states from text based on a multi-component emotion model", Computer Speech and Language Vol.36, 2016.*
56. *Anna Přibilová, "Spectrum Modification for Emotional Speech Synthesis", Multimodal Signals, LNAI 5398, Springer 2009.*
57. *Elliott D. Ross, "Affective prosody: What do comprehension errors tell us about hemispheric lateralization of emotions, sex and aging effects, and the role of cognitive appraisal" Neuropsychologia, Elsevier 2011.*
58. *Franco Orsucci, Roberta Petrosino, Giulia Paoloni, Luca Canestri, Elio Conte, Mario A Reda and Mario Fulcheri, "Prosody and synchronization in cognitive neuroscience" EPJ Nonlinear Biomedical Physics, Springer, 2013.*
59. *Jia-En Wang, "Emotional prosody perception and its association with pragmatic language in school-aged children with high-function autism", Research in developmental disabilities, Elsevier, 2015.*
60. *Steven B. Chin, Tonya R. Bergeson, Jennifer Phan, "Speech intelligibility and prosody production in children with cochlear implants", Journal of Communication Disorders, Vol.45, 2012.*
61. *Eszter Tislja´r-Szabo, Csaba Pleh, "A scribing emotions depending on pause length in native and foreign", language speech Communication, Vol.56, 2014.*
62. *Peerzada hamid ahmad, "Transformation of emotions using pitch as a parameter for Hindi speech", International Journal of Multidisciplinary Research, Vol.2 Issue-1, 2012.*
63. *S. S. Agrawa1, Nupur Prakash and Anurag Jain, "Transformation of emotion based on acoustic features of intonation patterns*

for Hindi speech", African Journal of Mathematics and Computer Science Research, Vol. 3(10), 2010.

64. *Geethashree.A and Dr. D.J Ravi, "Transformation of Emotions using Pitch as a Parameter for Kannada Speech", International Conference on Recent Trends in Signal Processing, Image Processing and VLSI, Association of Computer Electronics and Electrical Engineers, 2014.*
65. *Jiri Pribil and Anna Pribilova, "Evaluation of influence of spectral and prosodic features on GMM classification of Czech and Slovak emotional speech", EURASIP Journal on Audio, Speech, and Music Processing Springer 2013.*
66. *J. Pibil , A. Pibilová, "comparison of spectral and prosodic parameters of male and female emotional speech in czech and Slovak", Acoustics, speech and signal processing (ICASSP), international conference IEEE, 2011.*
67. *Manjare, Chandraprabha, Anil S. D., Shirbahadurkar, "Speech Modification for Prosody Conversion in Expressive Marathi Text-to-Speech Synthesis", International Conference on Signal Processing and Integrated Networks, IEEE, 2014.*
68. *Manjare Chandraprabha Anil S. D. Shirbahadurkar, "Expressive Speech Synthesis using Prosodic Modification for Marathi Language", 2nd International Conference on Signal Processing and Integrated Networks, IEEE 2015.*
69. *Ms. C.D.Pophale, Prof.J.S.Chitode, "Embedding Prosody into Neutral Speech", IOSR Journal of Electrical and Electronics Engineering,Vol.9, Issue 2, 2014.*
70. *Shashidhar G. Koolagudi, "IITKGP-SEHSC: Hindi speech corpus formation analysis", Devices and communications (ICDeCom), international conference, IEEE 2011.*
71. *Marc Schro¨ Der, "The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching", International Journal of Speech Technology, Vol.6, 2003.*
72. *Harpreet Kaur, "Prosody Modification of its output Speech Signal", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 5, 2014.*
73. *RashmiVerma, "Conversion of Neutral Speech to Storytelling Style Speech", Advances in pattern recognition (ICAPR), eighth international conference IEEE 2015.*